UNCLASSI

SECURITY CLAS

AD-A197 661

(2)

JMENTATION PAGE

1a. REPORT SE UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report # 12	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 88-0776
6a. NAME OF PERFORMING ORGANIZATION Texas A&M University	6b. OFFICE SYMBOL (If applicable) SD
6c. ADDRESS (City, State and ZIP Code) College Station, TX 77853	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research
	7b. ADDRESS (City, State and ZIP Code) Same as 8c
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) NM
8c. ADDRESS (City, State and ZIP Code) Bolling Air Force Base Washington, DC 20332	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-87-0149620-85-C-0144
	10. SOURCE OF FUNDING NOS.

BLG 410

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2304	A6	

11. TITLE (Include Security Classification)
COVARIANCE ANALYSIS IN GENERALIZED LINEAR MEASUREMENT ERROR MODELS

12. PERSONAL AUTHOR(S)
Carroll, Raymond J.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical Journal	FROM 8/87 TO 8/88		30

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	errors-in-variables, generalized linear models

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

We summarize some of the recent work on the errors-in-variables problem in generalized linear models. The focus is on covariance analysis, and in particular testing for and estimation of treatment effects. There is a considerable difference between the randomized and nonrandomized models when testing for an effect. For estimation, one is largely reduced to using an errors in variables analysis. Some of the possible methods are outlined and compared.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Major Brian Woodruff	22b. TELEPHONE NUMBER (Include Area Code) (202) 767-5026
	22c. OFFICE SYMBOL NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

COVARIANCE ANALYSIS IN GENERALIZED LINEAR MEASUREMENT ERROR MODELS

Technical Report # 12

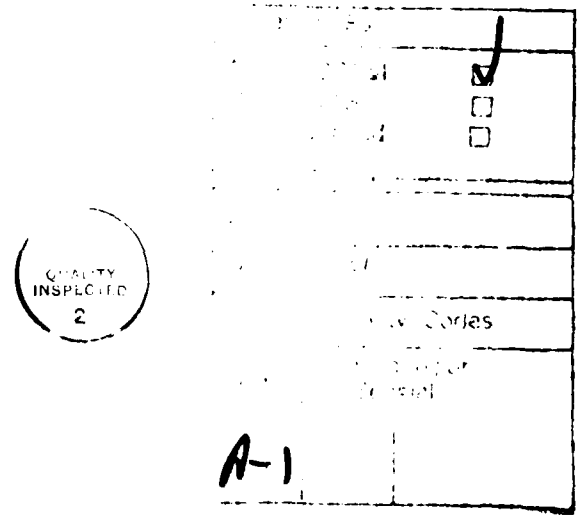
Raymond J. Carroll

COVARIANCE ANALYSIS IN GENERALIZED LINEAR MEASUREMENT ERROR MODELS

Raymond J. Carroll[1]
Department of Statistics
Texas A&M University
College Station, TX 77843

Abstract

We summarize some of the recent work on the errors-in-variables problem in generalized linear models. The focus is on covariance analysis, and in particular testing for and estimation of treatment effects. There is a considerable difference between the randomized and nonrandomized models when testing for an effect. For estimation, one is largely reduced to using an errors in variables analysis. Some of the possible methods are outlined and compared.



[1] This paper was prepared for the Workshop on Errors-in-Variables in Epidemiology at the National Cancer Institute, October 8-9, 1987, and it was supported by the Air Force Office of Scientific Research.

Acknowledgements

I thank Mitchell Gail and David Byar for inviting me to present an overview of measurement error models at the NCI workshop, and Len Stefanski, Cliff Spiegelman and Paul Gallo for many useful discussions. Rob Abbott introduced me to the problem in the context of the Framingham study, and most of my own work has been an attempt to answer the questions he posed.

Outline

Part I: Introduction and Overview

Section 1: Introduction

Section 2: Calibration and Validation Data Sets

Section 3: Outline

Part II: Simple Regression Models

Section 1: Linear Regression

Section 2: Deconvolution, Efficient Estimation and the Structural Model

Section 3: Other Simple Regression Models

Part III: Randomized Studies

Section 1: Covariance Analysis in Linear Regression

Section 2: Covariance Analysis in Logistic and Probit Regression

Section 3: Other Regression Models

Section 4: Non-covariance Models

Part IV: Nonrandomized Studies

Section 1: Covariance Analysis in Linear Regression

Section 2: Covariance Analysis in Other Regression Models

Part V: Discussion

Notation

Y = response variable

Δ = treatment assignment indicator

X = true value of the predictor

Z = observed value of the predictor

Part I: Introduction and Overview

1.1 *Introduction*

The aim of this paper is to survey approaches to estimation of and testing for treatment effects in the analysis of covariance when predictors (covariates) cannot be measured exactly. Because our primary focus is epidemiology, the discussion includes linear, logistic (success-failure), Poisson (counts) and survival time regression. As far as possible, we will avoid technical details and even formulae, refering the reader instead to the relevant literature. Some familiarity with statistical terminology is presumed, especially maximum likelihood estimation and weighted least squares.

When predictors are measured exactly as assumed in standard textbooks, it is sufficient to study simple regression, i.e., regression with one predictor. More complex models differ only in algebraic detail. When predictors are measured with error, studying only simple regression is a mistake. Some lessons from simple regression carry over to more complex models, but others do not. Covariance models are the simplest cases with generalizable lessons, so quite apart from their obvious importance in epidemiology they can be a central focus of study.

Synonyms for "predictors measured with error" are "errors-in-variables" and "measurement error models". By these phrases we mean that there is a concept of the true value of a predictor, but we can observe only an estimate of this true value. It is particularly important to note that the error in the estimate may have many components, not just a machine recording error. A recurrent theme is the importance of understanding and, often, modelling all the sources of error. We will return to this in part V.

A good example of a predictor measured with error is systolic blood pressure (SBP). One plausible definition of true SBP is a person's long term average reading from a 24-hour monitor. In practice, we observe only an estimate of this long term average. Sources of variability in the observed reading include not just simple machine error, but operator variability, conditions, time of day and even such subtle factors as to whether the patient is under (a short term period of) stress. Often, machine error can be made quite small, but even in a quality study Carroll, et al. (1984) estimate that 25% of population variability

in observed SBP is caused by other sources of error. In other studies in this volume, measurement error accounts for as much as 80% of the observed variability in predictors.

In epidemiology, it is important to distinguish between measurement error models and Berkson (instrumental variable) models. The former have been the focus of most of the research in the field and will be the main topic in this paper. Let X be the true value of a predictor and Z its observed proxy. In measurement error models, it is natural to model Z as a function of X. Very loosely, this means that we consider X to cause Z. For example, X might be true cigarette usage and Z reported usage, or X might be true radiation exposure and Z exposure as measured by a radiation badge. There are two subcases of the measurement error model. In the functional case, the true values of predictors (X) are considered fixed but unknown constants. This would be an appropriate framework if patients were gathered in some semi-deterministic way as opposed to being random samples from poulations, in which latter case X is random and we are in a structural case.

In the Berkson or instrumental variable model, it is most natural to model X as a function of Z. For example, in the paper by Tosteson, Schafer and Stefanski given in this volume, X is true NO_2 intake, which is most naturally modeled by a combination of NO_2 concentrations in the bedroom and kitchen. As a second example, Z might be the nominal amount of a drug given to a patient and X the true value; clearly the former determines the latter.

It is sometimes asserted that the whole problem of errors-in-variables is a waste of time. The argument goes something like this: "after all, you only get to observe Z, so you should be modeling the response Y as a function of Z rather than as a function of the true but unobservable covariable X". This glib and somewhat perverse viewpoint is dangerously ignorant. Here is one example of a measurement error problem which is important in epidemiology. It has been well known for at least twenty years that in an unbalanced linear analysis of covariance, measuring a covariable with error and then modeling the response as a function of the observed Z can lead to the WRONG conclusion. It is important to emphasize that we are talking about a totally invalid conclusion, not just an inefficient analysis. Other papers in this volume will make the same point in contexts other than covariance.

1.2 *Validation and Calibration Data Sets*

As other papers in this volume indicate, it is becoming increasingly common to have available a validation (or calibration) data set where both X and Z are observed. The idea is to use the validation data set to understand the error process, and then to transfer this knowledge across to the main data set of interest. It is especially with regard to this transfer that the distinction between the measurement error and Berkson (instrumental variable) models becomes vital. As we shall explore later, estimates such as maximum likelihood depend on knowledge of the distribution of the true X given the observed Z. In the Berkson case, it is by definition natural to model this conditional distribution, and in many cases this distribution will be the same across data sets.

For example, suppose X is true NO_2 intake, which is naturally modeled as a function of the combination Z of bedroom and kitchen NO_2. Here X is fundamentally a function of Z, and it is reasonable to suppose that the distribution of X depends only on the value of Z and not on the data set. The reader will find this to be the case in the paper by Tosteson, Schafer and Stefanski, where the fundamental relationship seems to hold even across continents.

In measurement error models, the fundamental relationship is that Z depends on X and not vice versa as in the Berkson case. It seems sensible to believe that given a fixed amount of true radiation exposure (X), the distribution of observed exposure (Z) should be constant from data set to data set. Thus, a validation data set gives information on the conditional distribution of observed given true. However, maximum likelihood and other estimates require the opposite conditioning, needing the distribution of true given observed. We can only transfer the knowledge of this conditional distribution from a validation data set in the measurement error context if the distribution of true exposures is the same in the validation and the main data set. This point may seem very subtle and of little consequence, but remember that if one transfers validation information inappropriately, one can get final estimates which are seriously in error.

For example, suppose that our validation data set measures true and badge recorded radiation exposure in a population with only moderate exposure, while our main data

set contains individuals who are heavily exposed. If one assumes that observed exposure depends only on true exposure, then the conditional distribution of true exposure given observed will differ in the two data sets. If we try to use this last conditional relationship in the main study, we can get badly biased results.

1.3 *Outline*

In Part II, we survey in depth the simple regression model. Parts III and IV discuss randomized and nonrandomized covariance models. Part V gives our conclusions and suggestions. Of particular interest is our conclusion with regard to measuring treatment effects in a generalized linear model with covariates which differ in the treatment and control groups. Large measurement error in such an unbalanced covariance problem can destroy the validity of the usual analysis based on ignoring the error. If measurement error is unavoidable, one must design the study so that one can understand the form and nature of this error.

Part II: Simple Regression Models

2.1 *Simple Linear Regression*

Measurement error models in linear regression is the subject of the authoritative book by Fuller (1987), and for this topic we will only reference those papers which we need directly. The classic simple linear regression model assumes that responses Y follow a linear curve in a true predictor X except for equation errors e:

$$Y = \beta_0 + \beta_1 X + e. \tag{2.1}$$

The equation errors have mean zero and standard deviation σ_e. Because of error, we observe Z, an estimate of X. Most papers assume that the error u in estimating X is linear:

$$Z = X + u, \tag{2.2}$$

where u has standard deviation σ_u. In either the functional or structural cases, if the true X's have mean μ_x and standard deviation σ_x, the ordinary least squares (OLS) estimates

of intercept β_0 and slope β_1 are biased, even for very large samples. The degree of bias depends on

$$\lambda = \sigma_x^2/(\sigma_x^2 + \sigma_u^2) = (1 + \sigma_u^2/\sigma_x^2)^{-1}. \tag{2.3}$$

In fact,

$$\begin{aligned} \hat{\beta}_0(\text{OLS}) &\to \beta_0 + \beta_1 \mu_x (1 - \lambda) = \beta_0^*; \\ \hat{\beta}_1(\text{OLS}) &\to \lambda \beta_1. \end{aligned} \tag{2.4}$$

Since λ is always less than or equal to 1.0, OLS underestimates β_1 in absolute value. This phenomenon is called ATTENUATION and is illustrated in Figure 1. Note that the OLS line correctly estimates the true line at the central value μ_x, but is in error elsewhere.

There are special features which deserve comment.

⊗ *Effect of Measurement Error on Bias:* The bias depends on the measurement error variability relative to the variability of X. Only when σ_u^2/σ_x^2 is large will OLS estimates be biased in important ways, see Fuller (1987, pp1-4).

⊗ *Effect of Prediction /Calibration:* Carroll and Spiegelman (1986) show that prediction and calibration confidence intervals are affected by the size of $(\beta_1 \sigma_u/\sigma_\epsilon)^2$. If this quantity is large (even when the bias in OLS is small), ignoring measurement error leads to confidence intervals which are much too conservative.

⊗ *Hypothesis tests for effect of X:* The usual test for an effect is $H_0 : \beta_1 = 0$. Equation (2.4) shows that under the null hypothesis, OLS estimates of β_1 are acceptable for use, and it turns out that standard hypothesis tests can be employed. This happy state of affairs does not carry over to more complex models, especially nonrandomized covariance models.

⊗ *Confidence Intervals for Slope:* For asymptotically valid 95% intervals for β_1, OLS will not work because of its bias. Even asymptotically valid confidence intervals require correction for measurement error. The difficulties in constructing exact finite sample confidence intervals are legion, see Gleser and Hwang (1985).

⊗ *Methods of Correcting Slope for Measurement Error:* If λ is known, then from (2.4) the obvious estimate is

$$\hat{\beta}_1(\lambda) = \hat{\beta}_1(OLS)/\lambda \tag{2.5}$$

Of course, λ is unknown and must be estimated. There is an enormous literature on methods for estimating λ, see Fuller (1987, Chapter 1). This is a daunting literature for the uninitiated, because methods depend either on artificial assumptions such as $\sigma_u = \sigma_\epsilon$ or on the use and type of replication on X or Y. In all cases, since $0 < \lambda < 1$, errors-in-variables estimates eliminate asymptotic bias but increase variability (tastes great - more filling!).

⊗ *Using Errors-in-Variables to Test for Slope Effects* To test $H_0 : \beta_1 = 0$, the increased variability of most errors-in-variables estimates results in less power than simply ignoring the measurement error. This magical but very special result is sometimes used (mistakenly) as a reason to ignore measurement error.

⊗ *Modelling the Measurement Error Process:* Just because most papers assume the additive error model (2.2) does not excuse one from designing the experiment so that it is possible to check this model. For example, suppose measurement error is multiplicative:

$$Z = Xu_*,$$

where u_* has mean 1.0 and standard deviation σ_{u*}. Then instead of (2.4) with λ given by (2.3), the OLS estimate of slope converges to

$$\hat{\beta}_1(\text{OLS}) \to \lambda_* \beta_1, \text{ where} \tag{2.6}$$
$$\lambda_* = \sigma_x^2/\{\sigma_x^2 + \sigma_{u*}^2(\mu_x^2 + \sigma_x^2)\}.$$

Clearly, if the measurement errors are multiplicative but one has pretended they are additive, then even "corrected" estimates of slope will be biased. This little fact is ignored in most of the literature; exceptions are Hwang (1986) and Stefanski (1985). We know of almost no literature which addresses the question: are the errors additive and homoscedastic? A doubting viewpoint is in Spiegelman (1986).

2.2 *Deconvolution, Efficient Estimation and the Structural Model*

Consider a structural case, i.e., the X's are random. Pretend that the measurement errors are additive as in (2.2) and normally distributed. Write the X's as having a distribution governed by a density function g. In this special case, various methods have evolved to obtain efficient estimates of β_0 and β_1. The first and most common (Fuller, 1987, Chapter 1) is to pretend that the X's are normally distributed and compute the resulting structural maximum likelihood estimate (SMLE), either assuming $\sigma_e = \sigma_u$ or by using replication. Gallo (1982) shows that the SMLE is of the form (2.5).

If the X's and e's are normally distributed, so too are the observed Z's. Although it may be done only rarely, one could plot a histogram of the observed Z's. What should one do if the resulting histogram is, say, very skew so that the X's are not normally distributed? The power of the SMLE is that it still can be used because it is of the form (2.5); estimates and tests are asymptotically valid even if the errors are not normally distributed. This is a remarkable model robustness property. The drawback is that without normality the SMLE is inefficient. We know of no way to improve efficiency without losing model robustness, at least in the sense of mathematical statistics. A practical method is to pretend that X has some sort of more general distribution, e.g., Weibull. We know of no instance where this device has been employed, although it has been anticipated by Schafer (1987a). Results of Carroll and Hall (1987) suggest that using the observed data to suggest a model for the X's may be impractical. In some sense, hypothesizing a flexible model for the X's is more guesswork than anything else. This is one instance where experimental design might be useful, i.e., devote part of the available resources to understanding the distribution of X.

Bickel and Ritov (1987) also try to improve upon the SMLE. They construct an estimator which makes this improvement as long as equation and measurement error are normally dsitributed. Unfortunately, the estimator is complex, relying on kernel density estimation techniques in its construction. Also, there is the (potentially delicate) problem of bandwidth selection in the kernel estimate. In addition, the role of the assumption of normally distributed equation and measurement error remains to be explored. These estimates are yet not ready for routine implementation.

There have been some attempts to construct alternatives to the SMLE which are robust against outliers. We believe that doing this will require some replication. For one early and partially successful attempt, see Carroll and Gallo (1982).

Perhaps the most interesting point here is that there are a lack of data diagnostic tools for checking assumptions, although see Schafer (1987b). Indeed, there is only a little work on the sensitivity of estimates to underlying assumptions. The SMLE is at least consistent with a standard theory assuming only additive errors, and hence forms a benchmark against which to make comparisons. Reasonable alternatives are robust estimators and maximum likelihood estimates without normality assumptions.

2.3 *Other Simple Regression Models*

Common models in epidemiology are logistic, Poisson, exponential and proportional hazards regression. The logisic model is used for failure-success data, the Poisson model for counts such as mutation, and the other two models are used to analyze survival times. There has been considerable recent interest in these models in the context of measurement error, but actual applications to data have been few.

These simple regression models share some of the features of simple linear regression. Typically (but not always!), the usual estimates are attentuated, see Stefanski and Carroll (1985). In testing for an effect due to X, i.e., $H_0 : \beta_1 = 0$, we can ignore measurement error. Unfortunately, estimation of β_1 is very difficult once one leaves simple linear regression.

The key feature that makes linear regression tractable is that the OLS estimate of β_1 estimates $\lambda\beta_1$ when λ is independent of β_0 and β_1, see (2.3) and (2.4). As in section 1.2, this leads to the natural corrected estimate $\hat{\beta}_1(\mathrm{OLS})/\hat{\lambda}$, which is asymptotically unbiased with a well-understood limiting normal theory for use in inference. Such easy corrected estimates are not available for other simple regression models.

Even for simple regression models, computing the maximum likelihood estimate (MLE) of β_1 requires specifying the distribution of X given Z (see Appendix A), although in proportional hazards regression Prentice (1982) indicates that we must specify the distribution of X given both Y and Z. The results of Carroll and Hall (1987) indicate that specifying such conditional distributions is difficult to do from information about only

Y and Z. It is useful to summarize some of the possible approaches to obtaining estimates of β_1. We consider only the case, as in logistic and Poisson regression, that it is sufficient to specify the distribution of X given Z.

In the rest of the section, we characterize seven possible approaches to obtaining estimates of the parameter β_1. These methods are as follows:

(1) Maximum likelihood for a specified parametric family;

(2) Quasilikelihood generalized least squares;

(3) Redesign the study to gather information on the distribution of X;

(4) Exploit special features of certain models;

(5) Approximate maximum likelihood by Taylor Series;

(6) Approximate quasilikelihood by Taylor series;

(7) Correct the usual methods by Taylor series

⊗ *Maximum Likelihood for a Specified Parametric Family.* If we can specify a parametic model for X given Z, then as in Appendix A we could perform a conditional likelihood analysis. This program has been carried out by Armstrong (1985) and Schafer (1987a). See section 1.2 for comments on the use of validation data sets to help specify such a model. In particular, note that validation data sets can be used for this purpose in the Berkson case but not for measurement error models unless it is reasonable to assume that the distribution of the true X is the same in the validation data as it is in the main data set of interest. For example, we might pretend that, given Z, X is normally distributed with mean $a + bZ$ and variance σ^2. Unfortunately, such an assumption is difficult to check from data on Z only, and violation of the normality assumption can degrade performance of the estimates, see Schafer (1987a) for a striking example. This method depends on X being a random variable. Perhaps more complex and flexible models could be used to lessen the dependence of the results on assumptions. Likelihood estimates in measurement error models do suffer from a distressing tendency towards occasional disasters, i.e., the estimates become wildly unrealistic. Even in the linear case, they have to be modified to have finite expectations, see Fuller (1980).

⊗ *Quasilikelihood Generalized Least Squares.* When a parametric family for X given Z can be specified, an attractive alternative to a likelihood analysis is based on quasilikelihood

generalized least squares (QGLS). These weighted least squares estimates are discussed by McCullagh and Nelder (1982), McCullagh (1983) and in Chapter 2 of Carroll and Ruppert (1988). As shown in Appendix A, specifying the distribution of X given Z enables us to compute the conditional mean and variance of Y given Z as a function of β_0, β_1 and a vector of nuisance parameters θ. A genuine advantage of QGLS estimates is that they are easier to compute than maximum likelihood estimates. Even in a fairly simple problem, Schafer (1987a) details the difficulty of maximum likelihood computations, while McCullagh and Nelder (1982) show that QGLS requires only iterative nonlinear least squares computations. A second advantage is that, being weighted least squares estimates, QGLS estimates have associated with them (1) routine inference based on asymptotic standards errors (2) graphical model checks; (3) simple model tests: and (4) influence diagnostics. These are surveyed in Chapter 2 of Carroll and Ruppert (1988).

Ease of computation and the ability to check some model assumptions make QGLS estimates an extremely attractive alternative. They seem to be restricted to the structural case.

⊗ *Exploit Special Features.* In logistic, Poisson and exponential models with additive normal measurement error, Stefanski and Carroll (1987a) show that it is unnecessary to specify the distribution of X given Z. Stefanski (1987) constructs estimates for β_1 in this case. His estimates share most of the defects of the Bickel and Ritov (1987) method discussed in the previous section. In the proportional hazards model, Prentice (1982) discusses a case that the measurement error, while large, is not crucial.

⊗ *Approximate the MLE by Taylor Series.* If the measurement error is "small", Whittemore and Keller (1986) show that an approximation to the MLE can be made by knowing the conditional mean and variance of X given Z. The resulting estimates had poor behavior in their Monte-Carlo study. Partly, this may be due to their use of mean squared error as a criterion for comparison. The experience from linear regression (Fuller, 1987) is that the MLE has infinite second moment, and in Monte-Carlo studies one will see a disaster in about 1% of the simulations. In the linear case a fix-up is available (Fuller, 1980), and this modified MLE does considerably better in simulations.

The crucial requirement for applying Whittemore and Keller's approximation is knowl-

edge of the conditional mean and variance of X given Z. While Carroll and Hall (1987) and Stefanski and Carroll (1987b) have shown that it can be very difficult to estimate the distribution of X given Z, Stefanski and Carroll (1988) have shown that the estimation of the moments of X given Z is considerably easier for additive measurement error, although still not simple.

The tendency for approximate MLE's to have occasional disasters limits their use. Until modified versions become available, these methods cannot be recommended.

⊗ *Approximate Quasilikelihood by Taylor Series.* Whittemore and Keller (1986) have approximated quasilikelihood estimates when measurement error is "small". Again, one must specify the mean and variance of X given Z. While this last is easier than specifying the entire conditional distribution, it is by no means easy to do from data. In their Monte-Carlo study, their approximate QGLS estimates performed quite well. The methods are appealing and worth further study.

⊗ *Correct the Usual Methods by Taylor Series.* This method has been developed in epidemiological models by Stefanski and Carroll (1985) and Stefanski (1985, 1987), while for classic nonlinear regression see Amemiya and Fuller (1985). The advantages of this method is that we need not specify or estimate the distribution of X given Z. Indeed, X need not be random, so that this method can be applied to the functional model. By Taylor series methods, one shows that the usual estimates of β_1 are estimating $\lambda\beta_1$ approximately for some λ. As in linear regression, one estimates λ and inverts. The method has worked well in the simulations of Schafer (1987a) and Whittemore and Keller (1986). The latter construct a simulation in which the moments of X given Z are known, which is advantageous to approximate quasilikelihood since it requires this knowledge. Even though the corrected estimates do not require these conditional moments, they are just as efficient as approximate quasilikelihood except in those rare cases that σ_x^2 is much smaller than σ_u^2.

An advantage of corrected estimates is that they can be easily applied, without assumptions about unobservables. Corrected estimates can serve as a diagnostic, indicating whether the effect of measurement error is substantial. The estimates, being approximations, are approximately consistent. Resampling methods can be used for inference.

Explicit formulae for corrected estimates are given in Appendix B.

Part III: Randomized Studies

3.1 *Covariance Analysis in Linear Regression*

Suppose that we have a control group, a treatment group, and a covariable X. For convenience, we consider completely random assignment to the two groups, there being no difference when more common randomization schemes employing blocking are used. Assignment is denoted by Δ, with $\Delta = -1/2$ and $\Delta = +1/2$ indicating assignment to the control and treated groups respectively. The response Y is linearly related to X with intercept depending on Δ:

$$Y(\Delta) = \beta_0 + \beta_1 \Delta + \beta_2 X + e.$$

The effect of treatment is β_1, so we wish to estimate and make inference about β_1. The question is: what happens when X is measured with error? Theoretical analysis is given by Carroll, Gallo and Gleser (1985) and Gleser, Carroll and Gallo (1987).

It may come as as major surprise that OLS ignoring measurement error gives (asymptotically) valid estimation and inference for the treatment effect, at least under a remarkably general set of conditions. The effect of the existence of measurement error is a decrease in the power of tests for β_1. Even so, OLS will most often be preferred to the *common* measurement error correction techniques, which are typically even less powerful.

Some intuition as to why OLS works for the treatment effect is given in Figure 2. Recall that the effect on measurement error on estimating a line is to attenuate the slope, although (like a see-saw) at the central value of X the value of the OLS line coincides with the correct value. Covariance analysis fits lines with the same slope to each group, with the treatment effect being the difference in the lines at the central value of all the X's (pooled over the two groups). Randomization insures balance: the central value of all the X's is the same as the central value of the X's in each group. This assurmes that OLS still estimates the correct treatment effect.

The effect of measurement error on OLS inference is a loss of power. In general, if there were no measurement error then the power for detecting a treatment effect in the

normal case is a function of the noncentrality parameter $N\beta_1^2/\sigma_\epsilon^2$, where N is the total sample size. With additive normal measurement error in X, the noncentrality parameter is smaller, being

$$N\beta_1^2/\{\sigma_\epsilon^2 + \beta_2^2(\sigma_z^{-2} + \sigma_u^{-2})^{-1}\}. \tag{3.1}$$

The loss of power can be substantial. It is important to note that the loss of power is related to the strength of the covariable as measured by β_2^2. An open problem is whether there are weak assumptions which enable one to construct tests with better power than OLS tests.

If the equation errors are normally distributed, and if the measurement error is additive and normally distributed, the Bickel and Ritov (1987) device could be used to get increased efficiency over OLS for inference about β_1, at least in principle. Practical implementation and the strong assumptions limit use of this alternative, at least at present.

When the X's are estimated by replicate Z's, the Carroll and Gallo (1982) robust estimates could be employed. Work in progress is improving on these methods.

3.2 Covariance Analysis in Logistic and Probit Regression

In logistic regression, the responses Y take on the values 0 and 1 representing failure or success. Given (x, Δ), the probability of success is

$$P\{Y = 1 \text{ given } X, \Delta\} = F(\beta_0 + \beta_1\Delta + \beta_2 X), \tag{3.2}$$

where F is the logistic distribution function

$$F(u) = \{1 + \exp(-u)\}^{-1}.$$

Probit regression satisfies (3.2) as well, but with F being the standard normal distribution function Φ. The ordinary method of estimation is maximum likelihood.

In linear regression with additive measurement error as in (2.2), if e, u and X are all normally distributed, then the true and observed data follow the models:

$$Y = \beta_0 + \beta_1\Delta + \beta_2 X + e;$$

$$Y = \beta_0^* + \beta_1\Delta + \beta_2^* Z + e^*.$$

The terms β_0^* and β_2^* are complex, while e^* is normally distributed. In other words, under these ideal conditions, the observed data from a linear covariance also follow a linear covariance with exactly the right treatment effect. This is another reason that OLS can be used for reference about β_1.

In binary regression, we do not have quite the same happy circumstances. One illustration of this is given by Carroll, et al. (1984). For probit regression with additive measurement error and both u and X being normally distributed, the true and observed probabilities of success follow these probit covariance models:

$$Pr\{Y = 1 \text{ given } \Delta, X\} = \Phi(\beta_0 + \beta_1 \Delta + \beta_2 X) \tag{3.3}$$

$$Pr\{Y = 1 \text{ given } \Delta, Z\} = \Phi(\beta_0^* + \lambda\beta_1 \Delta + \beta_2^* Z)$$

$$\lambda = \{1 + \beta_2^2(\sigma_u^{-2} + \sigma_x^{-2})^{-1}\}^{-1/2}.$$

These are ideal circumstances, but even so there are two important points to note. First, when there is no treatment effect ($\beta_1 = 0$), the observed data will also show no treatment effect. This suggests that we will be able to test for the presence or absence of a treatment effect without knowing anything about the form of the measurement error. Of course, there will be a loss of power compared to the case that X is measured exactly. The second consequence of (3.3) is less positive. Since $0 < \lambda < 1$, we see that if we use the observed data to estimate the treatment effect, then we are estimating $\lambda\beta_1$ and not β_1. The treatment effect is attenuated! To get an estimate of the true treatment effect β_1, one must estimate λ or equivalently β_2, i.e., one must do a proper errors-in-variable analysis. The methods for doing such an analysis are basically the same as outlined in Section 2.3. In Appendix B, we give a detailed treatment (with formulae) of corrected estimates for logistic regression with additive measurement error.

Measurement error reduces power in this example, and for that reason alone one will want to eliminate it. If measurement error has been controlled as best as possible, there is a temptation from the preceeding example to ignore it entirely. After all, by ignoring the errror we underestimate treatment effect and are being conservative. Whether attentuaion of treatment effect holds in general is unknown. Even if it does, it is not obvious to us that one should accept a severe underestimate of treatment effect.

The final point concerns testing for the treatment effect in the general case, i.e., logistic regression and other models for measurement error. One can easily show that, at least in theory, the usual hypothesis tests for $H_0 : \beta_1 = 0$ (such as the score test) have the wrong Type I error level. The major difficulty is that the observed data do not follow a logistic regression model (!). In our practice we have observed that the score test for treatment effect is usually adequate since the observed data still approximately follow a logistic model.

3.3 Other Regression Models

For generalized linear covariance models such as Poisson and exponential regression, the message is about the same as for logistic and probit regression. The usual score tests have the wrong level, but unlike in logistic regression the change from a nominal α can be considerable. A simple modified score test for treatment effect ($H_0 : \beta_1 = 0$) is available from Stefanski and Carroll (1988), and can be computed without modeling the measurement error process. Good estimates of the actual treatment effect requires a proper errors-in-variables analysis. There is one exception which is important in the Poisson and exponential cases. If we model the mean response in a multiplicative fashion as

$$E(Y) = \exp(\beta_0 + \beta_1 \Delta + \beta_2 X), \tag{3.4}$$

then the usual analysis consistently estimates the treatment effect β_1. This can be shown by applying the techniques of Gail, et al. (1984), the details appearing in Stefanski and Carroll (1988).

For exponentially distributed survival data, one can obtain confidence intervals for a treatment effect without knowing a model for the measurement errors. Suppose the survival times have mean (3.4). If we want to test $H_0 : \beta_1 = c$, it is sufficient to consider

$$Y^* = Y \exp(-c\Delta),$$

since Y^* follows the exponential model with mean

$$E(Y^*) = \exp(\beta_0 + (\beta_1 - c)\Delta + \beta_2 X).$$

Applying a modified score test for treatment effects to Y^* is the same as testing $H_0 : \beta_1 = c$. The set of all such c's which are accepted in a (say) 5% level test forms a 95% confidence interval for β_1.

We know of no work in this context for the proportional hazards model when covariables are measured with error.

3.4 Non-Covariance Models

In the covariance models we have studied, the treatment assignment Δ is measured without error and is unrelated to the covariable X measured with error. In non-covariance models, we might have predictors $\Delta_1, \ldots, \Delta_p$ measured without error and predictors $X_1, \ldots, X_r$ measured with error. As long as $\Delta_1, \ldots, \Delta_p$ are unrelated to $X_1, \ldots, X_r$, the lessons learned from our covariance model carry over virtually without change.

Alternatively, there are many instances where we are interested in testing whether any of the variables $(X_1, \ldots, X_r)$ measured with error are significant predictors. Some of the papers in this volume are devoted to this problem and various facets of it, and we refer the reader to these papers for extensive literature review. This is of course an easier problem because under the null hypothesis we have a model based strictly on the observables $\Delta_1, \ldots, \Delta_p$. See Tosteson and Tsiatis (1988) for a discussion.

Part IV: Nonrandomized Studies

4.1 *Covariance Analysis in Linear Regression*

We continue with the linear covariance model (3.1) but no longer assume that the study is randomized. Let μ_c and μ_T be the mean of X in the control and treated groups respectively; with equal allocation $\mu = (\mu_c + \mu_T)/2$ is the overall mean. For random assignment, $\mu_c = \mu_T$. In Figure 3 we plot a schematic of covariance analysis for an unbalanced study when X is measured without error and the treatment group has much larger values of the covariable than does the control group. In this picture, there is a negative effect due to the treatment. Figure 4 shows what happens to Figure 3 when X is measured with additive error; attenuation not only hides the true negative effect but causes an observed (illusory) positive effect. It is also possible for measurement error to

cause us to conclude that there is a treatment effect when it really does not exist.

Algebraically, suppose measurement error is additive as in (2.2). Let σ_x^2 be the sample variance of the true X in the pooled data and $d = \sigma_x^2 + \sigma_u^2 - (\mu_T - \mu_c)^2/4$. Then OLS estimates not β_1 but rather

$$\beta_1 + \beta_2 \sigma_u^2 (\mu_T - \mu_c)/2d.$$

This formula confirms Figure 4: measurement error in an unbalanced study can invalidate OLS estimation and inference. Factors affecting this are the degree of imbalance $(\mu_T - \mu_c)$, the measurement error (σ_u^2) and, not so surprisingly, the strength of the covariable (β_2).

In those cases for which OLS estimation and inference will not do, an errors-in-variables analysis is required. There are three main approaches: method of moments, functional estimates and structural likelihood analysis. For the former, one requires knowledge of the measurement error structure. For example, if an additive error model is appropriate, the OLS estimate of the vector $(\beta_0, \beta_1, \beta_2)^t = B$ estimates ΣB, where Σ is a matrix depending on the means and variances of X in each group as well as σ_u^2. If for each X we observe replicate X's, Σ can be estimated by $\hat{\Sigma}$ and the moments estimate is $\hat{\Sigma}^{-1}\hat{B}(\text{OLS})$, where $\hat{B}(\text{OLS})$ is the OLS estimate of B. In any case, the design should be such that Σ can be estimated. While estimates are easily constructed, asymptotic t-tests are not. With additive measurement error, formulae for standard errors are given by Fuller (1987). For multiplicative measurement error, see Hwang (1986).

In a functional analysis, each X is treated as a fixed constant to be estimated along with β_o, β_1, β_2 and σ_u^2. Assuming additive errors and replication of either X or Y or both, Fuller (1987) has constructed functional estimates of the vector $B = (\beta_o, \beta_1, \beta_2)^t$. These estimates are closely related to but more efficient than moments estimates and require no assumption of normal errors. Unfortunately, the functional approach appears to require additive errors.

The structural likelihood approach follows the same lines as the SMLE of section 2.1. Assuming additive measurement error, one pretends that this error follows a particular distribution, for example normal. For each group, one assumes that X also follows a particular distribution, say normal with means μ_c, μ_T and variances σ_c^2, σ_T^2 for the control and treated groups. Assuming some replication, a maximum likelihood analysis can be

performed.

To summarize, measurement error can wreak havoc on an unbalanced linear covariance analysis. To correct for the effects of additive measurement error, either a method of moments or functional approach is available with formulae for standard errors. For non-additive measurement error, formulae for estimates and standard errors are not generally available.

4.2 Covariance Analysis in Other Regression Models

Sufficiently severe measurement error in a predictively strong covariable can also invalidate an unbalanced, nonrandomized covariance analysis for logistic, probit, exponential or proportional hazards regression. See Stefanski and Carroll (1985) for a technical exposition in the logistic case. In the probit case with normally distributed measurement error and X normally distributed in each group, one can show as in Carroll, et al. (1984) that the usual estimate is approximately unbaised for a linear combination of the treatment effect β_1 and $\beta_2(\mu_T - \mu_c)$. While such calculations are useful in understanding those problems which are of greatest concern, the state of knowledge for these other covariance models is still primitive.

Possible approaches to an errors-in-variables analysis have been outlined in section 2.3. Formulae for corrected estimates in logistic regression with additive measurement error are given in Appendix B.

Proportional hazards regression does not operate exactly like the others. We believe that there may be cases, as in Prentice (1982), that imbalance in the covariables does not have a severe effect on treatment estimates. Work is in progress towards categorizing these cases.

Part V: Discussion

Regression models in epidemiology include not just linear but also logistic, probit, exponential and proportional hazards regression. The usual methods based on ignoring measurement error can give the wrong conclusions in an unbalanced, nonrandomized study of treatment effects. How bad the usual methods are depends in a complex way on the

size of the measurement error, the degree of imbalance in the study, how variable the true predictors are, and the strength of the covariable as a predictor.

Since correcting for measurement error is a difficult process, statisticians and epidemiologists must confront the issue in the design of a study, and not after a study has been completed. It is the equivalent of supporting motherhood and apple pie to suggest that we ought to try to eliminate measurement error whenever possible. The loss in power for detecting treatment effects even in a randomized study can be considerable when covariables are measured with significant error. Trying to correct for measurement error by statistical techniques rather than eliminating it in the design of the experiment is the same as the disastrously failed quality control strategy of inspecting for quality rather than building it in.

Unfortunately, we cannot eliminate measurement error in every instance, and a strategy has to be developed to cope with it. The first question to ask is: will it matter? This could be addressed through simple theoretical analysis. Although everyone should have access to a theoretical statistician (thus improving the already strong job market for same), a partial alternative is sensitivity analysis using simulation. By making reasonable guesses (based on pilot studies or experience) for the size of the measurement error, degree of imbalance, spread of the covariables and their strengths as predictors, one can simulate the experiment to see if a misleading conclusion is likely to be drawn. More importantly, evidence through simulation of loss of power or severe bias can act as a terrific incentive to careful study design.

If measurement error exists and is of likely importance, it has to be faced in the design of the experiment. Even the simplest errors-in-variables estimates (the approximate corrected estimate of Stefanski and Carroll (1985) for logistic regression and Stefanski (1985) for other models) require knowledge of the form of measurement errors and estimates of measurement error parameters. If we can safely assume additive measurement error with constant variance σ_u^2, replicate measurements (X's) have to be taken to estimate σ_u^2. Preferably, either a pilot study will have been performed or a proportion of study participants assessed exactly so that the critical assumption of additive homoscedastic measurement error can be checked.

If we had to choose an estimate based on available knowledge, we would favor the corrected estimates. These estimates are easy to compute under minimal assumptions on the measurement error process, assumptions which can be checked, at least partially, by replication. The available Monte-Carlo evidence is that corrected estimates do very well in practical situations.

Evidence may well be found to suggest that corrected estimates will not perform well in certain practical cases. In this instance, because of the tendency of maximum likelihood estimates to explode occasionally, for structural models where X is random in each group we prefer as alternatives to corrected estimates the method of quasilikelihood generalized least squares. The approximate version of QGLS developed by Whittemore and Keller (1986) is an intriguing possibility.

To compute a QGLS estimate, one is required to write a model for Y given the observed Z in terms of the original parameters. The most direct way to do this is to specify the distribution of the true X given both Z and treatment assignment Δ. If one takes this direct approach, three approaches will be useful immediately.

(1) *Write down a model.* For example, one might suppose that X given Z and Δ is normally distributed with mean $a + b\Delta + cZ$ and variance σ^2. The simulation of Schafer (1987a) shows that either a much more flexible model must be specified or the assumption checked. The latter can be done, for example, by measuring X exactly in a percentage of the study participants. Writing very flexible conditional models is a difficult task.

(2) *Construct indirect models.* It is sufficient to write a general model for X given Δ only and combine this with a model for Z given X. Operationally, this is often easier than the first method. Measurement error distributions are often simpler to characterize through pilot studies, and flexible models for X given only treatment assignment Δ are more readily available.

(3) *Measure X Exactly in Part of the Study.* If one measures X exactly in a proportion of the study participants, one will have some data to check model assumptions about X and the structure of measurement error. This will at least enable one to avoid disasters, e.g., assuming additive error when error is multiplicative or assuming normal X when

X has a very skew distribution.

In reviewing the literature, we have been struck by the overemphasis on the simple regression model and the corresponding lack of examples of techniques applied to data. As epidemiologists confront the issue of measurement error, we hope that the theoretical literature will evolve to solve the problems encountered in practice.

References

Amemiya, Y. and Fuller, W. A. (1985). Estimation for the nonlinear functional relationship. Preprint.

Bickel, P. J. and Ritov, Y. (1987). Efficient estimation in the linear errors in variables model. *Annals of Statistics*, 15, 513-540.

Armstrong, B. (1985). Measurement error in the generalized linear model. *Communications in Statistics, Series B*, 14, 529-544.

Carroll, R. J. and Gallo, P. P. (1982). Some aspects of robustness in functional errors in variables regression models. *Communications in Statistics, Part A*, 11, 2573-2585.

Carroll, R. J. and Gallo, P. P. (1984). Comparisons between maximum likelihood and method of moments in a linear errors in variables regression model. In *Design of Experiments: Ranking and Selection*, T. J. Santner and A. J. Tamhane (Eds). Marcel Dekker, New York.

Carroll, R. J., Gallo, P. P. and Gleser, L. J. (1985). Comparison of least squares and errors in variables regression with special reference to randomized analysis of covariance. *Journal of the American Statistical Association*, 80, 929-932.

Carroll, R. J. and Hall, P. (1987). Optimal rates of convergence for deconvolving a density. Preprint.

Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression.* Chapman and Hall, New York.

Carroll, R. J. and Spiegelman, C. H. (1986). The effect of small measurement error on precision instrument calibration. *Journal of Quality Technology*, 18, 170-173.

Carroll, R. J., Spiegelman, C. H., Lan, K. K., Bailey, K. T. and Abbott, R. D. (1984). On errors in variables for binary regression models. *Biometrika*, 71, 19-25.

Fuller, W. A. (1980). Properties of some estimators for the errors in variables model. *Annals of Statistics*, 8, 407-422.

Fuller, W. A. (1987). *Measurement Error Models.* John Wiley and Sons, New York.

Gail, M. H., Wieand, S. and Piantadosi, S. (1984). Biased estimates of treatment ef-

fect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, 71, 431-444.

Gallo, P. P. (1982). Consistency of regression estimates when some variables are measured with error. *Communications in Statistics, Series A*, 11, 973-983

Gleser, L. J. (1981). Estimation in a multivariate errors in variables regression model: large sample results. *Annals of Statistics*, 9, 24-44.

Gleser, L. J., Carroll, R. J. and Gallo, P. P. (1987). The limiting distribution of least squares in an errors in variables linear regression model. *Annals of Statistics*, 15, 220-233.

Gleser, L. J. and Hwang, J. T. (1985). The nonexistence of 100(1 - α)% confidence sets of finite expected diameter in errors in variables and related models. Preprint.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Annals of Statistics*, to appear.

Hwang, J. T. (1986). Multiplicative errors in variables models with applications to the recent data released by the U. S. Department of Energy. *Journal of the American Statistical Association*, 81, 680-688.

McCullagh, P. (1983). Quasi-likelihood functions. *Annals of Statistics*, 11, 59-67.

McCullagh. P. and Nelder, J. A. (1982). *Generalized Linear Models.* Chapman and Hall, London.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69, 331-342.

Schafer, D. W. (1987a). Covariate measurement error in generalized linear models. *Biometrika*, 74, 385-391.

Schafer, D. W. (1987b). Diagnostics for measurement errors in regression. *Journal of Business and Economic Statistics*, to appear.

Spiegelman, C. H. (1986). Two pitfalls of using standard regression diagnostics when both X and Y have measurement error. *The American Statistician*, 40, 245-248.

Stefanski, L. A. (1985). The effects of measurement error on parameter estimation. *Biometrika*, 724, 583-592.

Stefanski, L. A. (1987). Construction of optimal estimates in generalized linear measurement error models. Preprint.

Stefanski, L. A. and Carroll, R. J. (1985). Covariate measurement error in logistic regression. *Annals of Statistics*, 13, 1335-1351.

Stefanski, L. A. and Carroll, R. J. (1987a). Efficient scores and optimal scores in generalized linear measurement error models. *Biometrika*, to appear.

Stefanski, L. A. and Carroll, R. J. (1987b). Deconvoluting kernel density estimators.

Biometrika, to appear.

Stefanski, L. A. and Carroll, R. J. (1987c). Score tests in generalized linear measurement error models. Preprint.

Stefanski, L. A. and Carroll, R. J. (1988). Estimating functionals of unobservables in deconvolution problems. Preprint.

Tosteson, T. D. and Tsiatis, A. A. (1988). The asymptotic relative efficiency of score tests in the generalized linear model with surrogate covariates. *Biometrika*, to appear

Whittemore, A. S. and Keller, J. B. (1986). Approximations for errors in variables regression. Preprint.

Appendix A

Suppose that the density of Y given X is $f(y|x,\beta)$, the density of Z given X is $f(z|x,\theta)$, and the density of X is $f(x)$. The conditional density of X given Z is $f(x|z,\theta)$. The marginal density of Z is $f(z|\theta)$. The likelihood function of the observed data is

$$\int f(Y|x,\beta)f(Z|x,\theta)f(x)dx$$
$$= \int f(Y|x,\beta)f(x|Z,\theta)dx \quad f(Z|\theta)$$

The conditional likelihood for Y given Z is then

$$\text{(A.1)} \qquad \mathrm{L}(Y|Z,\beta,\theta) = \int f(Y|x,\beta)f(x|Z,\theta)dx,$$

while the k^{th} moment of Y given Z for use in a quasilikelihood analysis is

$$\text{(A.2)} \qquad \int Y^k \mathbf{L}(Y|Z,\beta,\theta)dy.$$

If θ is unknown, we can obtain a conditional maximum likelihood analysis using (A.1) or a quasilikelihood analysis using (A.2). Both require us to specify a parametric form for the density of X given Z.

Appendix B

A corrected estimator for logistic regression

For logistic regression, a corrected estimator was developed by Stefanski and Carroll (1985, section 3.1). This was generalized by Stefanski (1985, equation (2.10) and following) for generalized linear models. The idea is to correct the usual estimator for its bias. Suppose that we have a sample of size n and the i^{th} person has treatment assignment Δ_i, actual covariable X_i and observed covariable Z_i. Here X_i and Z_i could be vectors as would happen most often in practice. We assume additive measurement error with covariance matrix S_u. If some of the components of X_i are measured exactly, the corresponding elements of S_u are zero. Let $\hat{S}_u$ be an estimator of S_u, see below. Write

$$D_i = \begin{pmatrix} 1 \\ \Delta_i \\ Z_i \end{pmatrix} \text{ and } B = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

Let $\hat{B}$ be the usual estimator of B based on the observed data. The logistic distribution function is $F(u) = \{1 + \exp(-u)\}^{-1}$ with first and second derivatives $F^{(1)}$ and $F^{(2)}$ respectively. If $\hat{\beta}_2$ is the usual estimator of β_2, then the corrected estimate is

$$\hat{B}_c = \hat{B} + A^{-1}(J_1 + J_2), \text{ where}$$

$$A = \sum_{i=1}^{n} F^{(1)}(D_i^t \hat{B}) D_i D_i^t$$

$$J_1 = (1/2) \sum_{i=1}^{n} F^{(2)}(D_i^t \hat{B}) D_i (\hat{\beta}_2^t \hat{S}_u \hat{\beta}_2)$$

$$J_2 = \sum_{i=1}^{n} F^{(1)}(D_i^T \hat{B}) \hat{S}_u \hat{\beta}_2.$$

To implement this estimate, we require an estimate $\hat{S}_u$ of the measurement error covariance S_u. In some instances, S_u may be known from previous studies. In other cases, components of variance analysis may yield an estimate. For example, suppose that each person has m replicate observations of the covariable X. Then X_i will be the average of the m observed values. If S_i is the sample covariance matrix for the i^{th} person, then our estimate $\hat{S}_u$ is

$$\hat{S}_u = (mn)^{-1} \sum_{i=1}^{n} S_i.$$

including better estimates of standard errors. For making inference about $\beta_1 = \alpha^t B$ where $\alpha^t = (0 \quad 1 \quad 0)A$, one might use the percentile-t method based upon the approximate t

statistic for $H_0 : \beta_1 = c$ given by

$$T = \{\alpha^t \hat{B}_c - c\}/\{\alpha^t A^{-1} \alpha\}.$$

For references on bootstrap confidence intervals, see Hall (1988).

The problem of inference has not been addressed adequately in the literature. In the structural analysis of covariance framework, resampling techniques such as the bootstrap could be used. Let n_{-1} and n_{+1} be the sample sizes in the control and treatment groups respectively. Take a random sample of size n_{-1} with replacement from the control group and a sample of size n_{+1} with replacement from the treatment group. This forms a bootstrap sample and from it a bootstrap estimate $\hat{B}_c^t$ can be formed. This gives us access to the now voluminous bootstrap literature.

Figure 1: Attenuation due to measurement error

This schematic plot shows the effect of measurement error. The solid diamonds are the true data, fitted by the solid (OLS) line. The large circles are the observed data, obtained by interchanging the 1st and 4th X values and the 7th and 10th X values: the resulting attenuated (OLS) line is dashed.

Figure 2: Measurement error in a randomized linear covariance

The solid lines are the relationship between true covariable and the response, with the treatment the upper line and control the lower line. The dashed lines indicate the observed relationships after attenuation caused by measurement error. The treatment effect Δ is the difference between the lines, and is the same for both the true and observed data.

Figure 3: Unbalanced linear covariance for true X

Those values of the covariable greater than 6 represent the treatment group, while those less than 6 respresent the control group. The treatment effect from a linear covariance analysis is the difference between the two lines evaluated at 6 and respresented by the vertical line. The treatment group has lower responses having adjusted for the covariable.

Figure 4: Unbalanced linear covariance for observed Z

This is the observed version of Figure 3 after measurement error. Those values of the covariable greater than 6 represent the relationship between observed covariable Z and the response for the treatment group, while those less than 6 represent the same relationship for the control group. The treatment effect from a linear covariance analysis is the difference between the two lines evaluated at 6 and represented by the vertical line. In the observed data, the treatment group has higher responses having adjusted for the covariable.

Figure 1

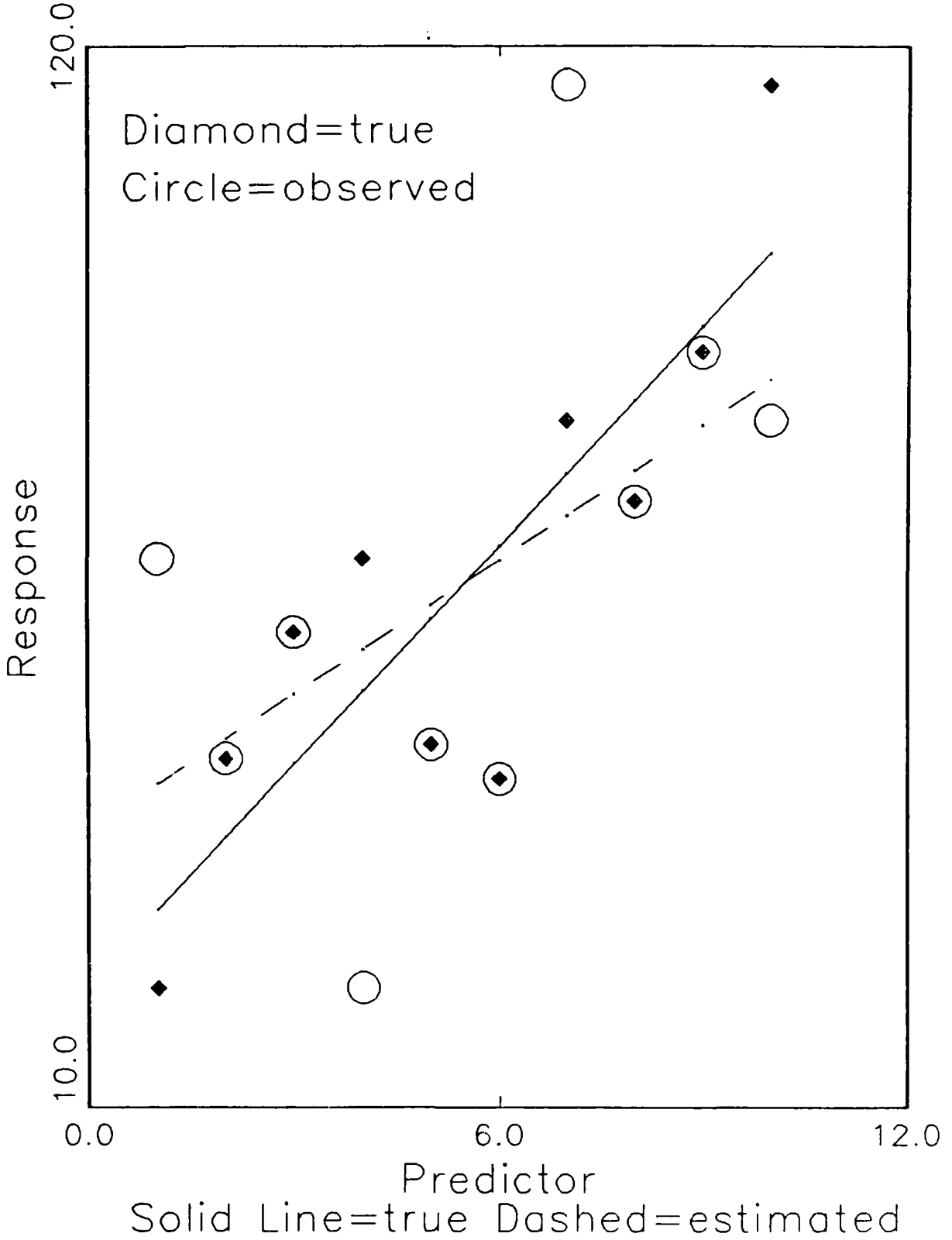

120.0
10.0
Diamond=true
Circle=observed
Response
0.0
6.0
12.0
Predictor
Solid Line=true Dashed=estimated

Figure 2

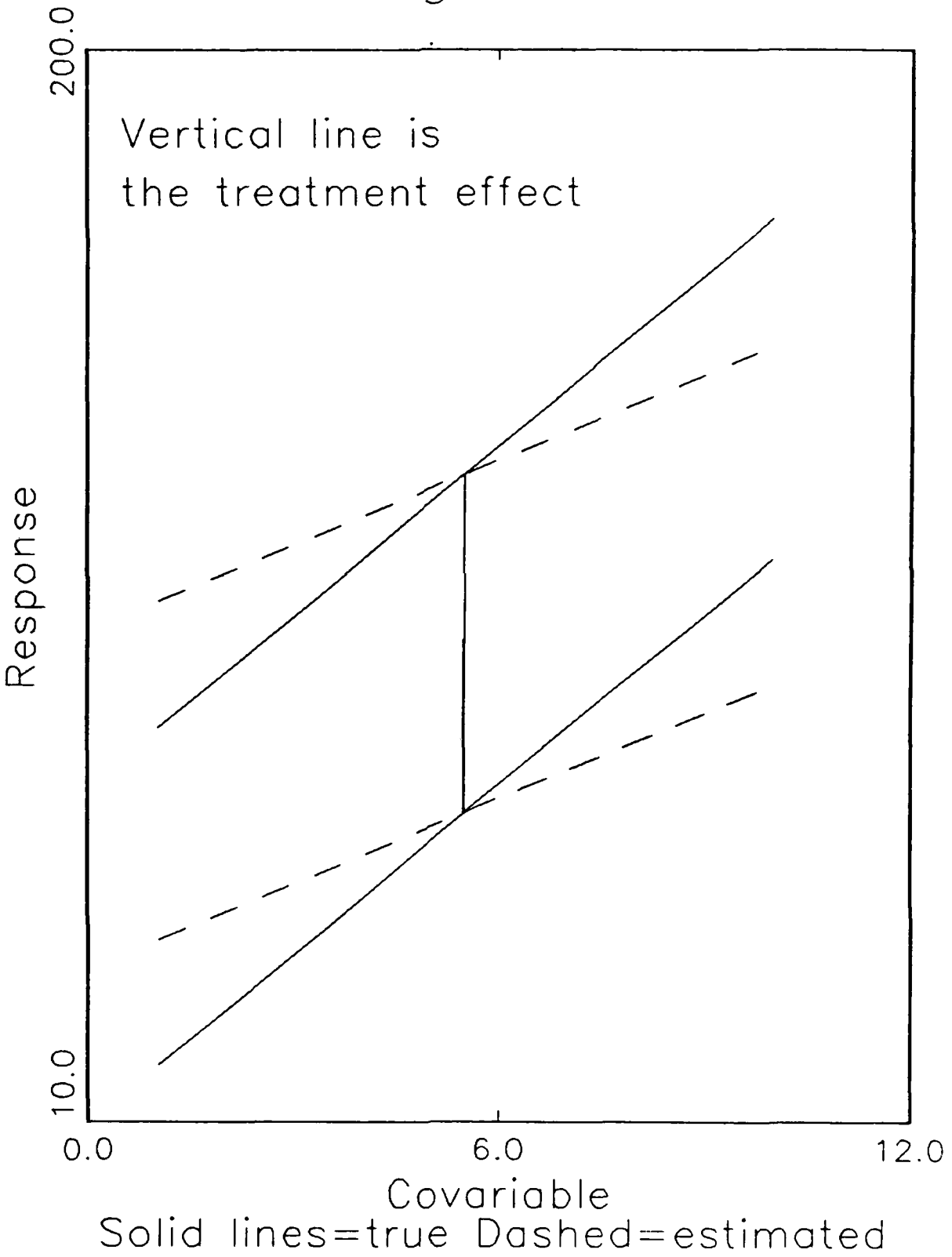

200.0
Vertical line is
the treatment effect
Response
10.0
0.0
6.0
12.0
Covariable
Solid lines=true Dashed=estimated

Figure 3

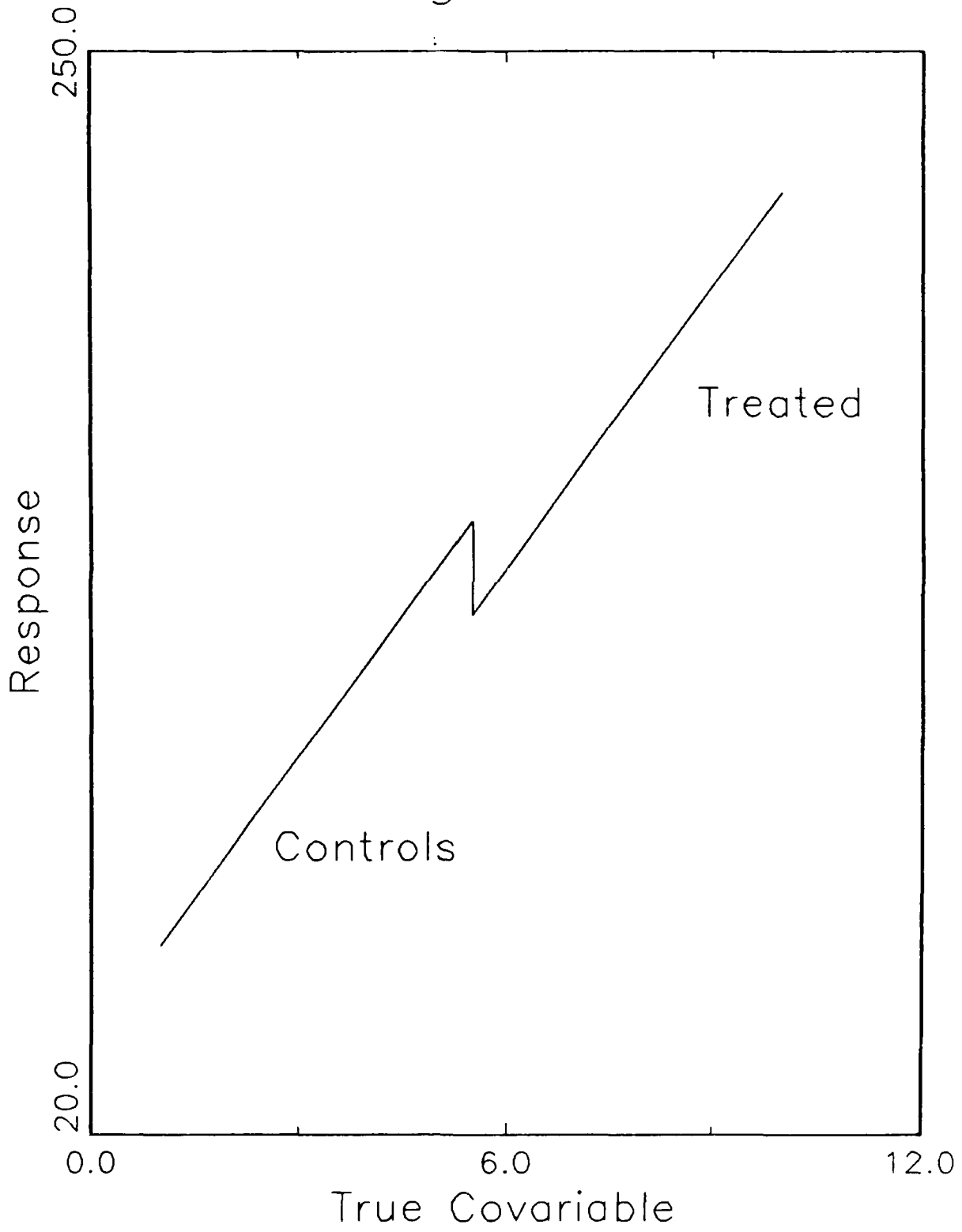
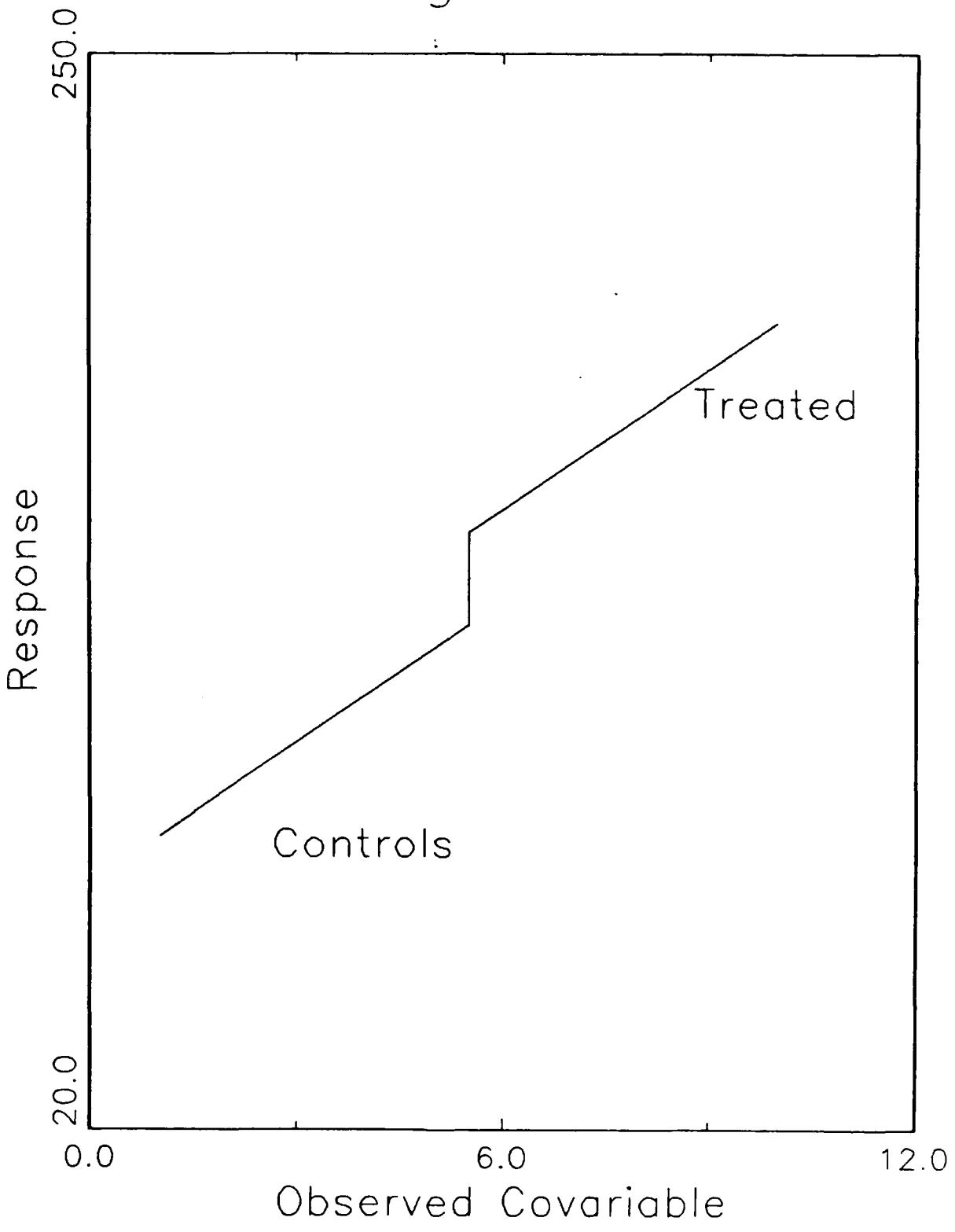

250.0
20.0
Response
Treated
Controls
0.0
6.0
12.0
True Covariable

Figure 4


250.0
20.0
Response
Treated
Controls
0.0
6.0
12.0
Observed Covariable